

# 基于知识发现的范例推理系统<sup>\*</sup>

倪志伟<sup>1,2</sup> 蔡庆生<sup>2</sup>

(安徽大学计算智能与信号处理教育部重点实验室 合肥230039)<sup>1</sup>

(中国科技大学计算机科学技术系 合肥230027)<sup>2</sup>

## Case-Based Reasoning System Based on Knowledge Discovery

NI Zhi-Wei<sup>1,2</sup> CAI Qing-Sheng<sup>2</sup>

(Ministry of Education Key Lab for Computing Intelligence and Signal Processing, Anhui University, Hefei 230039)<sup>1</sup>

(Department of Computer Science and Technology, University of Science & Technology of China, Hefei 230027)<sup>2</sup>

**Abstract** Nowadays the research and exploitation of the case-based system are getting more and more attention. Case-Based Reasoning (CBR) is a strategy for solving the object cases based on the source cases that are prompted by the object ones. CBR is not only a psychological theory for human knowledge, but will be a new cornerstone of the intelligent computer system technology. The case-based system is adopted in more and more application fields in order to obtain better results, especially in the fields with ill-defined and no expert knowledge. But there is a lot of knowledge required in CBR, and we are also faced with the same knowledge acquisition bottleneck as in the expert systems. Data Mining (DM) and Knowledge Discovery in Database (KDD) are just the most useful means to solve this kind of problem in order to make the knowledge acquisition more automated. In this paper, we discuss the data mining technology in CBR, especially we raise knowledge discovery in case base (KDC) and discuss this concept in detail. Finally, the structure of CBR based on DM is put forward.

**Keywords** Knowledge discovery in database, Data mining, Case base, Case-based reasoning

## 1 引言

范例推理(Case-Based Reasoning, CBR)是近十几年来人工智能中发展起来的区别于基于规则推理的一种推理模式,它是指借用旧的事例或经验来解决问题、评价解决方案、解释异常情况或理解新情况。CBR兴起的主要原因是传统的基于规则的系统存在诸多的缺点,如:在知识获取问题上存在困难,对于处理过的问题没有记忆而导致推理效率低下,不能有效地处理例外情况,整体性能较为脆弱等等,而 CBR 恰好能解决以上问题。

CBR 的显著优点有:信息的完全表达,增量式学习,形象思维的准确模拟,知识获取较为容易,求解效率高,等。然而,经过十多年的研究,人们也发现到 CBR 中存在一些问题,主要体现在以下几个方面:

(1) CBR 对噪音数据较为敏感,错误数据及冗余数据容易影响系统检索效率和求解效果;

(2) CBR 系统需要保持和管理一组数量较大的范例,时间和空间的复杂性都是必须仔细考虑的问题,否则有可能出现“范例库越大,系统性能越弱”的情况;

(3) 缺乏深层、表层背景知识以及不完整知识与由范例所表示的特殊知识的相互集成<sup>[1]</sup>;

(4) 范例工程过程的自动化问题,即范例知识的自动生成,如范例结构及其内容,相似性评估知识,现有范例库的自动更新,修正知识库的获取,索引模式,范例标识,等。而在这些知识的获取中也存在一定的瓶颈问题,如:修正知识的传统获取方法,范例库的获取,它们是领域专家与知识工程师通过不断积累而慢慢取得的。

对于以上问题,可以采取一些技术来解决。尤其对于(3)和(4),我们认为:基于已有机器学习技术<sup>[2]</sup>,在 CBR 中引入数据挖掘可以大大提高系统中知识获取的自动化程度。如在范例库中,引入知识发现技术,即范例库上的知识发现(Knowledge Discovery in Case base, KDC),它可以使得范例知识的获取高度自动化<sup>[3]</sup>,从而提高智能系统的整体性能。亦即,虽然范例推理系统比传统专家系统较为容易地获取知识,但范例推理系统也存在一定程度的知识获取瓶颈问题,如范例、修正知识、相似性评估等知识的获取问题<sup>[4]</sup>。我们通过在范例推理中引入数据挖掘技术,针对 CBR 系统中的多种数据源和多种知识,在其上进行数据挖掘,以提高知识获取的自动化程度<sup>[5]</sup>,加快智能系统的开发周期。

在 CBR 技术的研究中,国际上已有许多技术人员在探索 CBR 中的新技术和新方法。已有的成果有:Skalak<sup>[6]</sup>用遗传算法来选取范例的特征权值,一个用于 CBR 系统与决策支持系统的自动生成范例检索结构的机器学习方法<sup>[7]</sup>,David McSherry<sup>[8]</sup>主要讨论了基于需求的 CBR 中的递归修正技术,它采用的是针对预测领域的递归技术。关于范例库的自动获取,也有文献发表<sup>[9]</sup>,该文叙述了从象棋大师的棋谱数据库中自动获取范例的过程。

又如,NEC 公司研制的 Case-Method<sup>[10]</sup>,Rostock 大学研制的能预测传染病发展情况的 CBR 与 DM 混合系统<sup>[11]</sup>,微软使用了 Bayes 技术研制出“Aladdin”<sup>[12]</sup>,这是一个软件检测的 CBR 诊断系统,Salford 大学开发了一个 Bayes 网络数据挖掘系统以索引 CBR 系统<sup>[13]</sup>,一个称为 INBANCA<sup>[14]</sup>的系统是用来模拟一个足球比赛环境,其中的技术也是用 Bayes 方法。

<sup>\*</sup> 本课题得到国家自然科学基金重大项目资助(90104030)。倪志伟 博士,副教授,研究方向:机器学习、知识发现。蔡庆生 教授,博士生导师,研究方向:机器学习、知识发现、协调式人工智能。

倪志伟 博士,副教授,研究方向:机器学习、知识发现。蔡庆生 教授,博士生导师



从以上可以看出,国内外关于 CBR 与 DM 相结合的研究只是处于起步阶段,研究不够深入,也不系统。本文将系统地讨论在 CBR 中可以进行的数据挖掘技术,包括在 CBR 中基于数据库的数据挖掘技术,以及基于范例库的数据挖掘技术,本文研究的重点在于后者。

## 2 范例推理中基于数据库的数据挖掘技术

在数据库上进行辅助 CBR 方面的工作主要有二点:自动知识发现,范例知识的维护。首先是建造相关知识结构,这涉及范例工程的初始状态与 CBR 系统运行时的知识获取;其次是一个现有范例知识结构的优化与维护。具体地,涉及三个方面的内容:

(1)从数据库中建立范例一级知识结构 范例的识别以及它的表示、定义、结构都是很重要的。有三类数据挖掘技术可用来从数据库中建立范例库结构,即聚类、粗糙集、主成分分析。这些方法都可以用来发现定义范例结构的特征项。与此相关的一个重要问题是:特征项的权值。从数据库中可以挖掘出范例库中的特征项的权值。确定特征项权值的方法包括:信息论方法、神经网络、遗传算法。另外,还可以搜索数据库来补充范例的相关信息,如 Gibbs 样条法可用来发现范例数据中的空缺值。

其次就是相似性的表示与定义。一般地,使用距离度量来估计相似性,这些方法的缺点在于它们都没有考虑到领域的特殊知识。而数据库是一个很强大的领域特殊知识存储库,正好可以使用数据挖掘来生成相似性评估的知识。使用数据挖掘技术来建立相似性的评测方法有模糊技术、统计方法。

再者, CBR 中的修正过程是不可缺少的,修正知识库是关键性的知识。有大约十种不同的修正方法可以使用。其中在数据库中可以使用的数据挖掘技术有:规则发现、分类、聚类、遗传算法等。

(2)从数据库中建立范例库的组织结构 在范例库中,范例需要很好地组织以便有很高的搜索效率,这就是所谓的索引问题,且这些结构和处理是动态变化的,以便系统能不断地学习,此即系统的动态记忆能力。在一给定数据集上,运用数据挖掘技术有助于以多种方式来自动获取索引结构。首先,知识丰富的范例库组织策略常用语义网,这种表达的方式能将表层与深层知识紧密地结合起来。数据挖掘中的分类与聚类方法提供了一个适当的机制,以生成其中的概念与关系,并充分表达基于语义网的范例结构。其次,在范例库中将范例以合适的索引方式组织起来,对于一个给定数据集的自动分类,这种带有范例知识概念与层次聚类方法的紧耦合的一般领域知识是非常实用的。再次,在 CBR 系统中用到原型概念。在数据挖掘中,原型是分类和聚类的结果,即分类和聚类可以对范例索引知识的自动获取很有帮助。

(3)从数据库中更新范例知识的结构 自动更新范例库的过程涉及到范例库与知识结构的优化和维护。这个阶段不同于范例知识生成阶段。其中一个重要的突破在于一个称为“沼泽问题”的解决方法<sup>[15]</sup>,它是通过使用大型数据库作为 CBR 的存储来达到,其解决办法是只有对维护整个系统的能力有益的范例保持在范例库中。显然,一些数据挖掘技术可以用来确定数据库中的记录,这些记录对范例库的性能有益,数据挖掘的方法有分类、聚类,即寻找普通情况下的原型范例和特殊情况下的离群值。借助于辅助知识(如修正知识、相似矩阵、索引知识)来更新范例库,有益于提高 CBR 系统的能

力。

## 3 范例库上知识发现的主要步骤与目标

范例库是范例推理系统中最重要的知识库,它存放已有的经验或历史事实。在范例库上进行的知识发现,我们称为 KDC (Knowledge Discovery in Case base)。KDC 类似于 KDD,但又不同于 KDD 技术<sup>[16]</sup>。

范例库上知识发现过程的主要步骤是:

(1)范例库的维护:在对范例库进行数据挖掘之前,必须先对范例库实施维护操作,以删除范例库中的噪音数据和错误数据,尽量保证所挖掘出来的知识可靠有效。

(2)确定目标:根据要求,确定要解决的问题的目的。

(3)确定数据挖掘算法:根据所确定的任务,选择合适的知识发现算法。

(4)数据挖掘过程:这是 KDC 技术中的核心部分,也是关键技术。其主要任务是确定发现知识的种类,研究数据挖掘的模型及算法,发现规则知识。

(5)模式解释:对发现的模式进行解释说明。

(6)知识评价:发掘出的知识是否有价值,还要经过分析评价才能确定,最终得到有用的知识。

KDD 与 KDC 有许多类似的地方,也各有差别。两者的相同之处在于:

(1)KDD 与 KDC 的基本过程相似,许多数据挖掘的技术都可以有共同的原理。

(2)当范例库用关系数据库来表示时,许多 KDD 的技术与方法可以用在 KDC 之中。在目前的 CBR 系统开发过程中,许多有上万条范例的范例库均用数据库管理系统(DBMS)来管理,由于 DBMS 的成熟、通用、方便,所带来的 CBR 系统具有较好的软件工程开发环境与应用前景。

KDD 与 KDC 的不同之处在于:

(1)范例推理中的范例表达方式有多种形式,如语义网、数据库、框架、原型、神经网络等。而 KDD 的操作仅限于对数据库进行运算。

(2)在 KDD 中,数据的准备阶段比 KDC 中的数据预处理阶段要复杂得多,因为范例库中的数据是经过一定的处理而得到的。

(3)KDC 是面向 CBR 系统的,它是在 CBR 中的各个阶段使用的,并用于范例推理过程之中;而 KDD 可以面向许多系统,它在许多领域都可以应用,如在管理信息系统、地理信息系统、决策支持系统、专家系统中,等等。

范例库上知识发现过程的主要目标有:

(1)范例库的简约与特征提取:简约是采用图表、关联规则和可视化等技术得到范例子集的紧致描述;特征提取则是从范例库的一组子集中提取出关于这些数据的特征式,它们表达了范例子集的一般特征。

(2)挖掘范例库中特征项的权重:在范例推理系统中,常用的相似性判断是特征项之间的比较。相似性有表象相似与本质相似之分。如何表达范例间的本质相似,必须要找出范例库中特征项的权重。利用数据挖掘技术从范例库中发现其权重,无疑比人为设置的权重合理得多且正确得多。

(3)发现领域知识:可以从范例库中发现领域知识,并以函数、规则、图的形式表示出来,它们可以用在检索、修正等推理步骤中,如在检索阶段可以识别出有用的特征项和不重要的特征项,以及在学习阶段来解释范例。

(4)发现修正知识:CBR 中的修正过程是不可缺少的,修正知识库是关键性的知识库之一。修正阶段是 CBR 推理过程中的难点,主要是因为缺少获取和利用修正规则的方法学。目前的学习修正知识的方法有:

①利用领域知识来学习修正规则,此方法的优点是它可以与其它基于知识的系统相互利用领域知识;

②交互式修正规则的学习,一般是从专家用户处学习修正知识;

③从范例库中学习修正规则,一般采用归纳比较技术,即检查范例之间的差异来自动地学习到一批修正规则。这其中的问题有如下几点:如何选择范例来比较,如何在检索与修正间进行最大程度地合作,如何求精得到的规则集,如何应用修正规则,等等。

(5)范例库的学习:在 CBR 中,系统的学习体现在范例库及一些知识库的增长过程中,主要是新的范例不断地加入到范例库里。当新输入的问题通过 CBR 系统解决以后,是否加入到范例库中去,这可以通过多种方法来判断决定,常用的技术是在范例库中找出近似于新问题的旧有问题,以此决定学习与否;或从范例库挖掘出多个相似范例的聚类,抽象出一个范例保存在范例库中,以及与已有范例的合并等。

(6)范例推理系统的维护、范例库的维护:CBR 中的维护包括各种知识和结构的维护,最重要的是范例库的维护。范例库维护最重要的问题是:如何防止范例库在学习过程中的无限增大<sup>[16]</sup>,其解决办法一般是过滤掉对整个系统的能力无益的范例,即删除法。显然,一些数据挖掘技术可以用来确定范例库中的范例对范例库的系统性能有多大贡献,来确定它是保留还是删除,用数据挖掘的方法对范例库进行分类或聚类操作,即寻找普通情况下的原型范例和特殊情况下的离群值,并借助于辅助知识来更新范例库,如修正知识、相似性度量、索引知识等,以此来维护 CBR 系统。范例库的维护还包括冗余范例的删除、错误范例的寻找以及删除等。

#### 4 范例库上知识发现的主要方法与技术

在范例库上,知识发现过程中可以采用的具体挖掘技术和方法有:

(1)神经网络(NN) 它是一种通过训练来学习的非线性预测模型<sup>[3,17]</sup>,可以完成分类、聚类、范例特征挖掘等任务。用在 CBR 中的常用模型有:前馈网络,反馈式网络,自组织网络,相互激活与竞争网络,自适应共振网络,等。如在 CBR 中,范例匹配常采用自适应共振理论(ART)神经网络。ART 网络是一个模式分类器,具有学习新的训练模式,且又不会使以往的训练消失的特点,非常适合知识的获取与扩充。当然,ART 与基于反向传播算法的多层前馈神经网络(BP 网络)的结合,可以实现动态聚类与从聚类模板到每一范例的索引,可以大大提高范例的检索效率。可以利用 BP 网络建立一种 CBR 的检索模型,这个模型的最大特点是检索的速度与库的大小呈非线性关系,即范例库的大小不直接与检索速度相关,克服了基于比较的检索模型中检索速度与库的大小存在着线性关系这个缺点。另外,通过对范例库的训练与学习,利用从训练好的网络中提取规则的算法,我们可以从范例库中提取到一般性的规则形成知识库,以简化范例库。

为了判定范例间的相似度,我们以三层 BP 网为例来叙述。构造神经网络的目的,就是要通过它获取专家关于判定范例相似的知识。输入层节点由两个范例的特征项作为输入,输出层只有一个节点,表示范例的相似度。在用 BP 模型时,将范例库中各个范例与目标范例作为输入,计算其相似度,然后进行排序,实际使用时应该设定阈值,相似度大于阈值的范例

都可以被选取作为相似范例的候选。

(2)粗糙集 它不需要预先给定某些属性的数量描述,而是直接从给定问题的描述集合出发,通过不可分辨关系和不可分辨类确定给定问题的近似域,从而找出该问题的内在规律<sup>[18]</sup>。如,将粗糙集理论引入 CBR 中,充分利用范例库中冗余属性的简化,形成范例的索引,从而可以根据不同问题按不同索引进行检索并得出结论。粗糙集还可以用来过滤掉已有 CBR 中的不相关特征项。粗糙集也可以用在历史数据库的数据挖掘中,以寻找范例库中的特征项,这是从历史数据库直接生成范例库的步骤之一。此外,粗糙集与模糊技术的结合也可用在 CBR 系统之中。

我们用粗糙集来进行范例检索的操作,大致的思路如下:

①对于输入的问题,检查其核属性的值是否确切,如果不确定,则说明条件不充分,无结果,结束;否则,找到包含于输入问题的已知属性集的范例索引;

②对于求得的范例索引中的每个属性,根据输入问题的对应属性的取值,判断其归属于基于该属性所划分的等价类中的哪一个类;

③求出关键属性的等价类的交集,如果为空,则说明不存在具有确定性解的相似范例,只能交给专家用户去处理,并将其添加到范例库中,结束;

④否则,求出该交集与其它属性的等价类的交集,如果为空集,则说明在这个属性集中没有相似范例,然后重新选择范例索引;否则交集即为输入范例的相似范例集,并从基于决策属性所划分的等价类集中,找到包含相似范例集的等价类,此描述即为输入范例的解。

(3)基于规则和决策树的方法 采用规则发现和决策树技术来发现数据模式的核心是用某种归纳算法。通常是先对范例库中的范例进行操作,产生规则和决策树,再对它们进行分析和预测。例如,用归纳法可以完成范例推理中的范例检索。它是利用基于决策树的学习算法来实现的,从范例的各个组成部分抽取最能将该范例与其它范例区分开的成分,并根据这些成分将范例组织成一个类似于判别网络的层次结构,检索时采用判别树搜索策略。

(4)CBR 技术 作为数据挖掘的方法之一,CBR 方法本身也可以用来挖掘范例库中的知识,例如,通过建立修正范例的范例库,来记录以前范例修正的过程与经验,这种方法的特点是增加范例库的维护开销,交互时间的代价较高,但是实现的难度不大。

(5)遗传算法 它是属于一种基于生物进化和遗传学机理的概率搜索方法,群体搜索策略和个体之间的信息交换是它的主要特点。遗传算法可以在范例库中寻找一般规则,遗传算法还可以用在范例库中冗余范例的发现上。遗传算法也可以用来完成特征项赋权空间的直接搜索<sup>[9]</sup>。关键是 k-NN 算法的协作;它被用来作为评估函数用在人口遗传中每个染色体的适应度函数上。在实验中,染色体包含许多基因,每个基因是二维编码,代表了单个特征项权重,组成染色体的基因数量等同于在实例集中独立特征项的数量。

(6)数理统计方法 使用这种技术一般是首先建立一个数学模型或统计模型,然后根据模型提取出有关的知识。如用多元回归分析来学习 CBR 中每个特征项的权重。又如,可由训练数据建立一个 Bayes 网络,根据网络的一些参数和联系权重提取出相关的知识<sup>[19]</sup>。Bayes 方法可用于范例库的索引,也可用于从范例库中选取最相似的范例;用 Bayes 模型修正已找出的范例的解,以适应新范例的情形;用 Bayes 网络可以从范例库中提取出所需要的知识;也可以用 Bayes 来进行 CBR 学习,更新 Bayes 网络以增加新范例到范例库中,等等。

(7)关联规则挖掘 挖掘关联规则就是发现存在于范例库中的相关性,这可以用于规则的泛化或知识的提取。比如,在范例库中挖掘出关联规则,形成泛化规则,以减少范例库中的范例,这是范例库维护的技术之一。

(8)支持向量机(Support Vector Machine,SVM) 是近几年发展起来的新型通用的知识发现方法,在分类方面具有良好的性能。SVM 是建立在计算学习理论的结构风险最小化原则之上,其主要思想是针对两类分类问题,在高维空间中寻找一个超平面作为两类的分割,以保证最小的分类错误率<sup>[20]</sup>。

用 SVM 实现分类,主要是从原始空间中抽取特征,将原始空间中的样本映射为高维特征空间的一个向量,将其转化为优化问题,以解决原始空间中线性不可分的问题。

将 SVM 应用于 CBR 系统之中,这是个前沿方向及具有挑战性的课题。

## 5 基于知识发现技术的 CBR 系统框架结构

我们完整地设计了一个基于知识发现技术的 CBR 系统,见图1。具体功能包括有:从大型数据库中获取范例库,从数据库中学习到修正知识,监视整个 CBR 系统的性能,记录下失败的范例,不断更新和修正 CBR 系统中一些知识,使得系统的性能得以不断完善。此系统的初始状态是:有历史数据库、空范例库、空修正知识库。系统最终要达到的目的是:通过多次的重复循环,找到包含有整个问题空间的核心范例库,一组协调一致的修正知识。

重要的问题在于如何平衡操作数据库(Operational Data Base, ODB)、范例库以及修正知识库之间的种种关系。ODB 是整个系统的输入,它包括了自动收集的数据、范例的表示形式可以是类似于关系数据库的记录。范例建造器(Case Constructor)是使用 DM 技术(如粗糙集,模糊集等)从数据库中提取出范例库;数据挖掘引擎(Data Mining Engine)用来从范例库中挖掘出修正知识,并从 ODB 中补充一些遗漏的修正知识;领域知识库(Domain Knowledge Base)包含了 CBR 引擎中用在检索范例的领域知识;当向 CBR 系统提供一个输入问题时,CBR 引擎(CBR Engine)即可完成整个范例推理的操作;CBR 功能监视器(CBR Competence Monitor)使用来自于 CBR 引擎的反馈信息,并将这些信息存储在加强型操作数据库(Enhanced Operational Database, EODB)中,EODB 包含有 ODB、所遇到的关于系统功能的信息以及用来提高系统功能的失败范例。

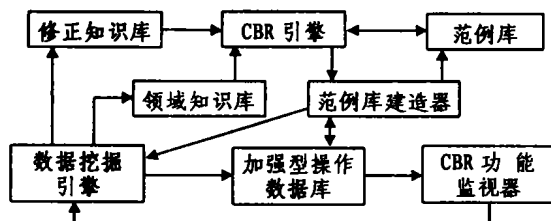


图1 CBR 的完整应用系统框架图

根据这个完整的框架,我们在实现时,可以简化这个基于知识发现技术的 CBR 系统。比如,历史数据库是以前该领域的运行数据记载,它是 CBR 系统的基础数据之一。在历史数据库的基础之上,我们可以利用知识发现过程来挖掘出范例推理过程所必需的范例库;整个系统的推理过程依赖于 CBR 的推理方式,由 CBR 推理机制来完成,范例库的维护可以由它来控制。修正知识库既可以来源于人机界面,又可以从历史

数据库和范例库中自动生成;知识发现模块控制整个系统的数据挖掘过程,包含了各种实现的数据挖掘方法与技术。

结束语 本文主要系统地研究了在范例推理系统中可以使用的数据挖掘技术,其一,从历史数据库中进行知识发现,有:从数据库中建立范例一级知识结构,从数据库中建立范例库的组织结构,从数据库中更新范例知识的结构。其二,特别提出一类新的概念及其范畴,即从范例库中进行知识发现(KDC)技术,包括有:特征项权重的发现,修正规则的挖掘,范例库维护中冗余范例的发现,为范例库的约简而作的数据挖掘处理等。接着对 KDC 的主要步骤、目标、方法等作了全面的研究。本文最后提出了一个完整的 CBR 应用系统模型,并给出了一个简化的基于知识发现技术的 CBR 系统实现过程,为本文的后续研究奠定了基础。

## 参考文献

- Jagielska I, Matthews C, Whitfort T. An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems. *Neurocomputing*, 1999, 24: 37~54
- Tsukimoto H. Extracting rules from neural networks. *IEEE Transactions on Neural Networks*, 2000, 11(2): 377~389
- Kraslawski A, Pedrycz W, Nystrom L. Fuzzy neural network as instance generator for case-based reasoning system. *Neural Computing & Applications*, 1999, 8: 106~113
- Azuaje F, et al. Discovering relevance knowledge in data: a growing cell structure approach. *IEEE Transactions on systems, man, and cybernetics*, 2000, 30(3)
- Anand S S, et al. Discovering case knowledge using data mining. In: *Proc. of PAKDD-98*, 1998. 25~35
- Skalak D. Prototype and features selection by sampling and random mutation hill-climbing algorithm. In: *Proc. 11th Int. Machine Learning Conf.*, Morgan Kaufmann, 1994. 293~301
- Babaka O, Whar S Y. Case-based reasoning and decision support systems. In: *IEEE Intl. Conf. on Intelligent Processing Systems*. Beijing, 1997. 1532~1536
- McSherry D. Demand-driven discovery of adaptation knowledge. In: *Intl. Joint Conf. on Artificial Intelligence*, Volume 1, 1999. 222~227
- Flinter S, Keane M T. On the automatic generation of case libraries by chunking chess games. In: *Proc. of the First Intl. Conf. on Case-Based Reasoning*, 1995. 421~430
- Kitano H, Shimazu H, Shibata A. Case-Method: A methodology for building large-scale case-based systems. In: *Proc. of AAAI-93*, 1993. 303~308
- Bull M, Kundt G, Gierl L. Discovering of health risks and case-based forecasting of epidemics in a health surveillance system. In: *Jan Komorowski and Jan Zytkow, editors, Principles of Data Mining and Knowledge Discovery*, 1997. 68~77
- Sbrees J, Heckerman D. Decision-theoretic case-based reasoning. In: *Proc. of the Fifth Intl. Workshop on Artificial Intelligence and Statistics*, 1995. 56~63
- Rodriguez A F, Vadera S, Sucar L E. A probabilistic model for case-based reasoning. In: *I Smith, B Faltings, eds. Cased-Based Reasoning Research and Development, Proc. of ICCBR-97*, 1997. 623~632
- Aha D W, Chang L W. Cooperative Bayesian and case-based reasoning for solving multiagent planning tasks. [Technical report]. Navy Center for Applied Research in AI, Naval Research Laboratory, Washington, DC, USA, 1996
- Smyth B, Keane M T. Remembering to forget: A Competence-Preserving Case Deletion Policy for Case-Based Reasoning Systems. In: *Proc. 14th Intl. Joint Conf. on Artificial Intelligence*, 1995. 337~382
- Stephan R. Knowledge discovery in scientific data. In: *Proc. of the Intl. Society for Optical Engineering*, 2000. 250~258
- Yahia M E, et al. Rough neural expert systems. *Expert systems with applications*, 2000, 18: 87~99
- Hu X H. Knowledge discovery in databases: an attribute-oriented rough set approach. [Ph. D. Thesis]. Computer Science Faculty of Graduate Studies, University of Regina, 1995
- Cheeseman P, Stutz J. Bayesian classification: theory and result. In: *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1996. 153~180
- Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 1997